Rsik Assessment of NewYork City Roads for Bicycle Accidents

Nidish Murugan - Siddhesh Suresh Kadam

2024-08-16

Executive Summary

Cyclist safety in New York City (NYC) has become an increasingly urgent concern as urban cycling gains popularity. This project addresses this challenge by developing a predictive model that assesses the risk of bicycle collisions on city roads. The model specifically focuses on non-behavioral factors, such as road conditions and infrastructure deficiencies, which are often overlooked in traditional safety analyses. By leveraging a comprehensive collection of datasets from NYC's Open Data portal, the project identifies high-risk routes and offers insights that could guide policymakers on where to focus safety interventions.

The model was built using Random Forest regression, a robust machine learning technique, and tested across multiple configurations. The final model, with an optimal number of variables randomly selected as 236, demonstrated strong predictive performance with a Root Mean Square Error (RMSE) of 0.0864, a Mean Absolute Error (MAE) of 0.0610, and an R-squared value of 0.4080.

These results indicate that the model can effectively identify high-risk routes for cyclists, capturing a significant portion of the factors influencing collision risk. The findings reveal that strategic improvements in road conditions and infrastructure could significantly reduce the frequency and severity of bicycle collisions. The precision and reliability of the model make it a valuable tool for enhancing urban cycling safety, providing actionable insights for both cyclists and city planners to mitigate risks and foster a safer environment for urban cyclists.

Objective

The primary objective of this project is to develop a predictive model that assesses the risk of bicycle collisions on various routes in New York City, focusing on non-behavioral factors such as road conditions and infrastructure. By identifying and analyzing these risk factors, the project aims to provide actionable insights that can help cyclists choose safer routes and support city planners in implementing targeted safety interventions. Ultimately, the goal is to enhance cyclist safety in NYC by shifting from reactive to proactive safety measures, reducing the incidence of bicycle collisions across the city.

Introduction

Urban cycling has become an increasingly popular mode of transportation in New York City due to its sustainability, cost-effectiveness, and ability to navigate through traffic congestion. The rise in cycling activity is driven by environmental awareness, healthier lifestyle choices, and the city's efforts to expand bike lanes and improve infrastructure. However, this growth has also led to a significant increase in bicycle-related accidents. The city's dense and complex road network, high traffic volumes, and interactions between cyclists, vehicles, and pedestrians, particularly at intersections, pose substantial safety challenges.

Traditionally, safety measures in New York City have been reactive, addressing areas with a history of accidents. While these interventions are necessary, they often fail to anticipate new high-risk areas that emerge due to changes in traffic patterns and urban development. This reactive approach overlooks the dynamic nature of the city's transportation environment, where ongoing construction and infrastructure changes can introduce new hazards. A more proactive strategy is needed to predict and address potential risks before accidents occur.

This project aims to shift from reactive to proactive safety strategies by predicting high-risk routes for cyclists, allowing for preventive measures to be implemented. The focus is on analyzing non-behavioral factors such as road quality, lane markings, and the presence of cycling infrastructure, which are within the control of city authorities. By addressing these physical and environmental factors, the project seeks to create a safer environment for cyclists, reducing the likelihood of accidents and improving overall safety in New York City.

Motivation

New York City continues to witness a high number of bicycle collisions, particularly at intersections and during peak traffic hours. Current safety measures are often reactive rather than proactive, focusing primarily on areas with a history of accidents. There is a pressing need for a datadriven approach to predict high-risk routes, enabling preventive actions to reduce collisions and enhance cyclist safety. By examining road-related factors that contribute to bicycle collisions, this project seeks to shift the focus towards more effective, proactive safety

About Data Source

The foundation of this project is built on a rich dataset sourced from the NYC Open Data portal, which provides extensive information on various aspects of traffic and transportation in the city. The primary datasets used in this study include:

1. Motor Vehicle Collisions - Crashes: This dataset contains over 2.1 million records of crash incidents, offering detailed information on crash dates, times, locations, and contributing factors. It serves as the core dataset for understanding where and why bicycle collisions occur.
2. Motor Vehicle Collisions - Vehicles: With over 4.2 million records, this dataset provides crucial details about the types of vehicles involved in collisions, their states of registration, travel directions, and other factors that might influence the dynamics of accidents.

Data Sources

- The data was sourced from New York City's Open Data portal, accessible at https://opendata.cityofnewyork.us/.
- The dataset is available for download in a ZIP file attached in the DropBox.

Data Handling

The format of the date feature is formatted properly to maintain consistency and since year and month are one of the main features for the bicycle accidents happening. Those features are extracted from the Data Feature and introduced as new Features.

Followed by Filtering the bicycle accidents. Since, the dataset has all accidents from all types of vehicles. The dataset is filtered where cyclist injured and killed are greater than 0

There are features in the data set such as '*VEHICLE.TYPE.CODE.1*', '*VEHICLE.TYPE.CODE.2*', '*VEHICLE.TYPE.CODE.3*' which has details on what kind of vehicle got into accdient. So, all the different types of bicycles and spelling errors are Standardized for consistency.

Setting the CRS to 4326 and removing all the NULL value Latitude and Longitude

Exploratory Data Analysis

Monthly Bicycle Crashes from 2012 - 2023

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

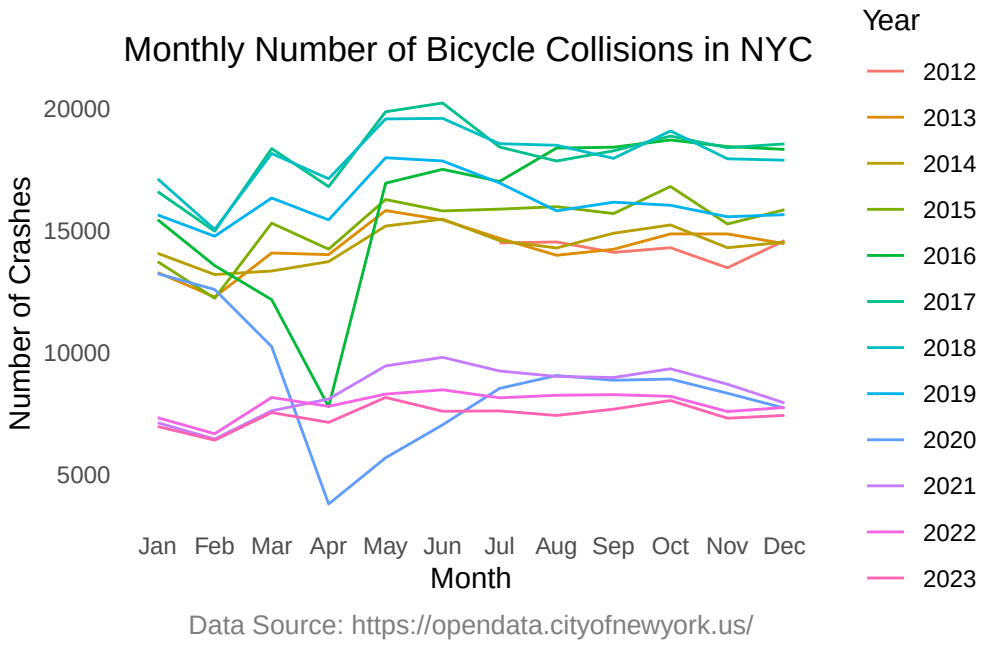

Figure 1: Monthly Trend of Bicycle Crashes in NYC

Figure 1 shows the monthly trends in bicycle collisions in New York City from 2012 to 2023. Each line represents a different year, illustrating the number of bicycle crashes that occurred each month.

Key Observations:

- *Consistency in Trends*: Bicycle collision trends have remained fairly stable year to year, with only minor variations. This indicates that the underlying risk factors contributing to bicycle crashes are

relatively stable over time.

- *2020 Dip*: The noticeable dip in 2020, likely due to COVID-19, introduces beneficial randomness into the data. This irregularity can improve the robustness of predictive models by helping them account for unexpected events and fluctuations.

Explanation for Using the Last Five Years of Data:

- *Relevance and Recency*: The last five years of data are likely the most relevant for understanding current cycling trends and the latest developments in road infrastructure. Given that the trends have been fairly consistent over time, using the most recent data captures the latest patterns while minimizing the impact of older data that might not reflect current conditions.
- *Impact of Randomness*: The irregularity observed during 2020 due to COVID-19 introduces valuable randomness into the data, which can enhance the predictive power of risk assessment models. By including this data, the models become better equipped to handle unusual patterns or anomalies, which are essential for assessing risk in a dynamic urban environment like New York City.
- *Computational Efficiency*: Focusing on the last five years also reduces the computational power required to process large datasets, making the analysis more efficient while still capturing the essential trends and variations needed for accurate predictions.

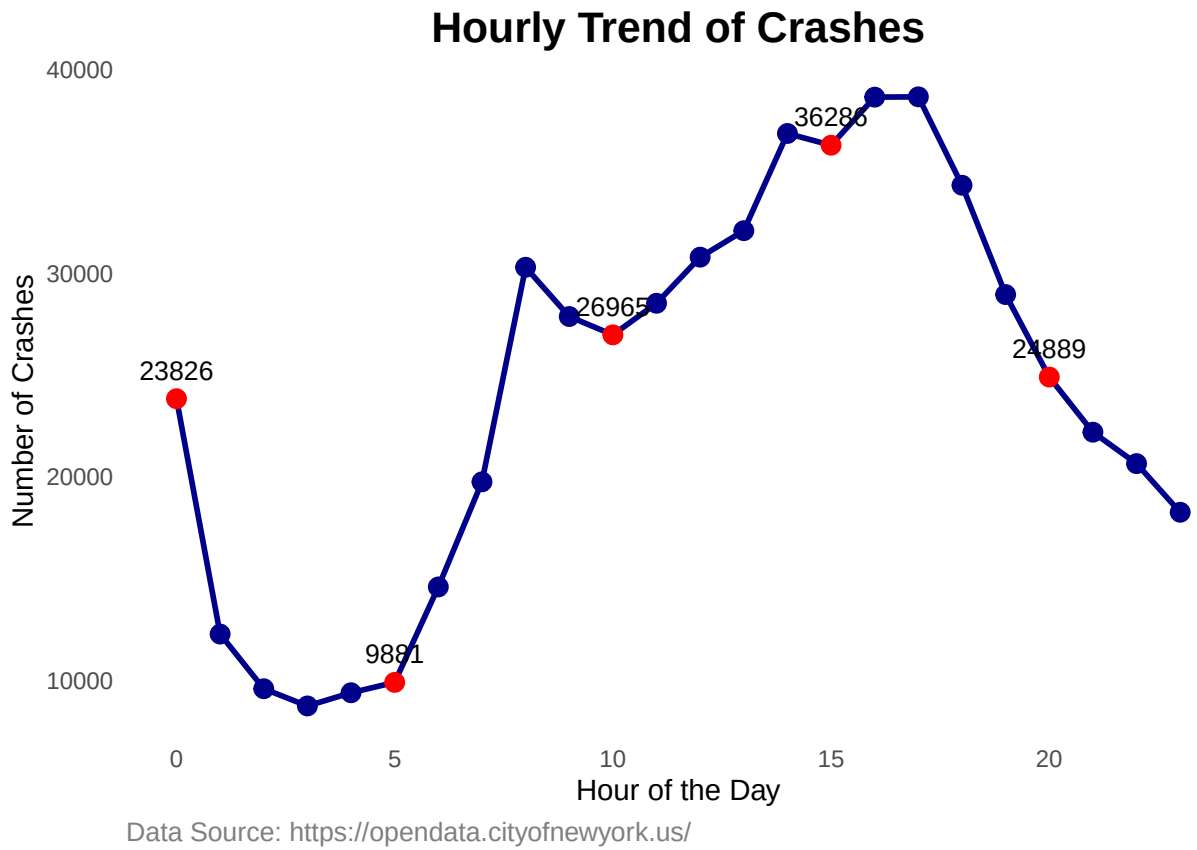

Figure 2: Hourly Trend of Bicycle Crashes in NYC

Peak Crash Hours: The highest number of bicycle crashes occur around 10 AM and 3 PM, aligning with increased traffic during rush hours and mid-afternoon. This suggests targeted safety interventions during these times could be effective.

Implications: Focused Interventions: Addressing crash peaks during morning and afternoon rush hours with enhanced safety measures could reduce accidents. Off-Peak Safety: Maintaining safety during low-traffic hours through better lighting and road adjustments is crucial. This highlights the need to consider time-specific strategies for reducing bicycle crashes.

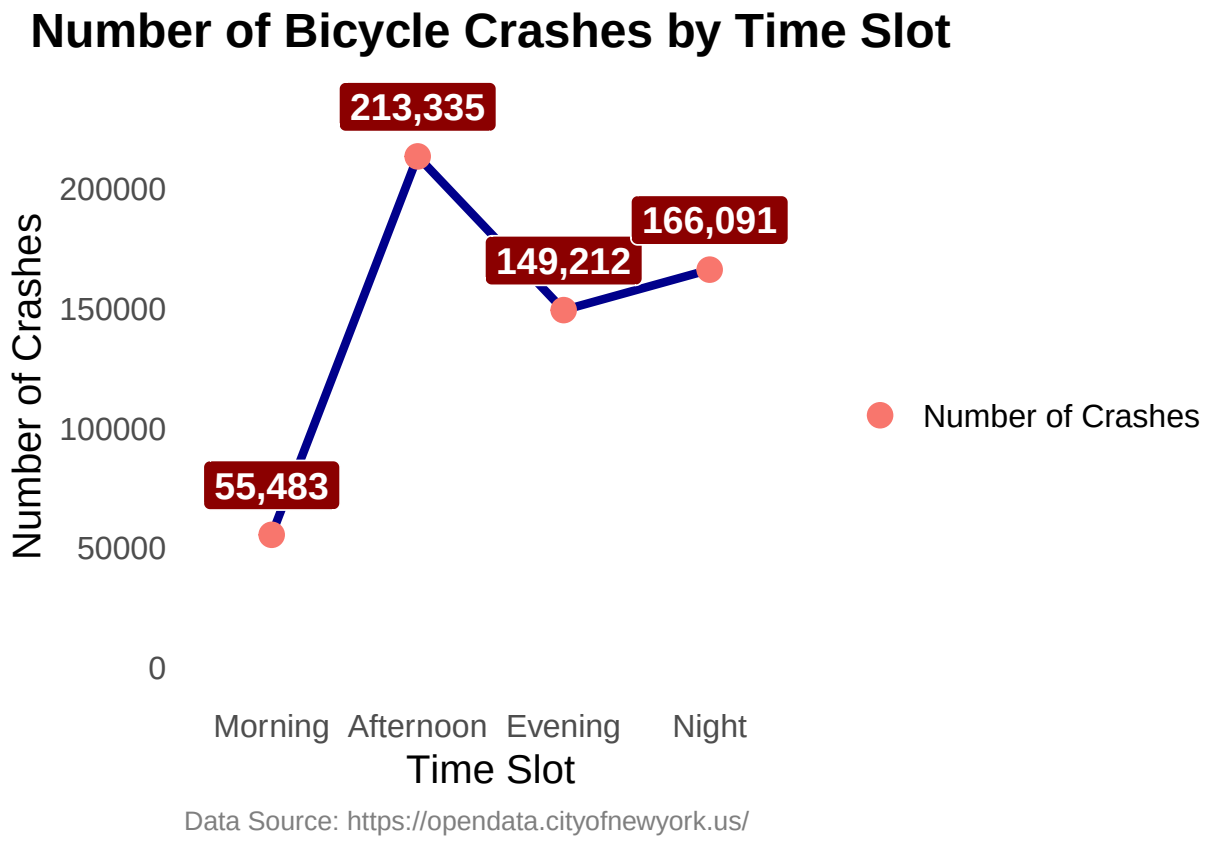

Figure 3: Number of Bicycle Crashes by Time Slot in NYC

Analysis

Afternoon Peak: Both analyses from Figure 2 and Figure 3 confirm the afternoon as the most dangerous period for bicycle crashes, driven by high traffic.

Explanation:

The plot shows the distribution of bicycle accidents across various ZIP codes in New York City (NYC) for the last five years (2019 - 2023). The intensity of the color represents the number of bicycle accidents, with darker colors indicating higher accident counts. ### Key Observations: - *High-Risk Areas*: The plot clearly shows that certain areas, especially in Manhattan and Brooklyn, have much higher bicycle accident rates. These are busy, high-traffic neighborhoods where cyclists face greater risks.

- *Improving Safety*: By pinpointing where most accidents occur, city planners can focus on making those areas safer for cyclists. This might include adding protected bike lanes and improving road conditions, ultimately encouraging more people to bike safely across the city.

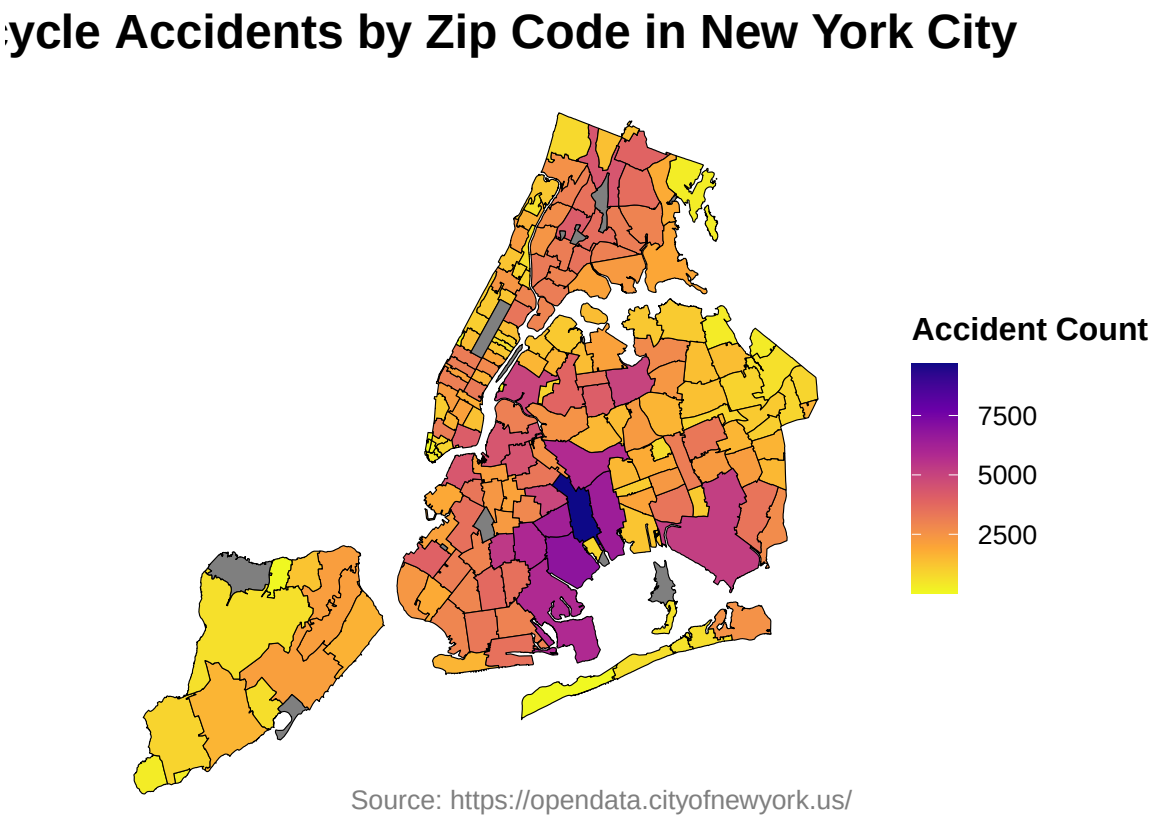

Figure 4: Displaying collision hotspots across NYC by highlighting areas with a high frequency of accidents from 2019 - 2023

Focusing on Manhattan, one of the boroughs of New York City, allows us to manage the computational intensity associated with extracting and plotting detailed road data for the entire city. Manhattan, being a densely populated and highly trafficked area, provides a significant and representative sample of the road and crash dynamics present in NYC. By concentrating on this borough, we can conduct a thorough analysis of Non-Behavioral factors while keeping the computational load manageable. This approach ensures that we can perform detailed spatial analysis without the risk of overwhelming computational resources, which would be necessary if we were to extend the same level of detail to the entire city.

```
## Data (c) OpenStreetMap contributors, ODbL 1.0. https://www.openstreetmap.org/copyright
```

```
## Some legend labels were too wide. These labels have been resized to 0.59, 0.39, 0.46. Increase legend
```

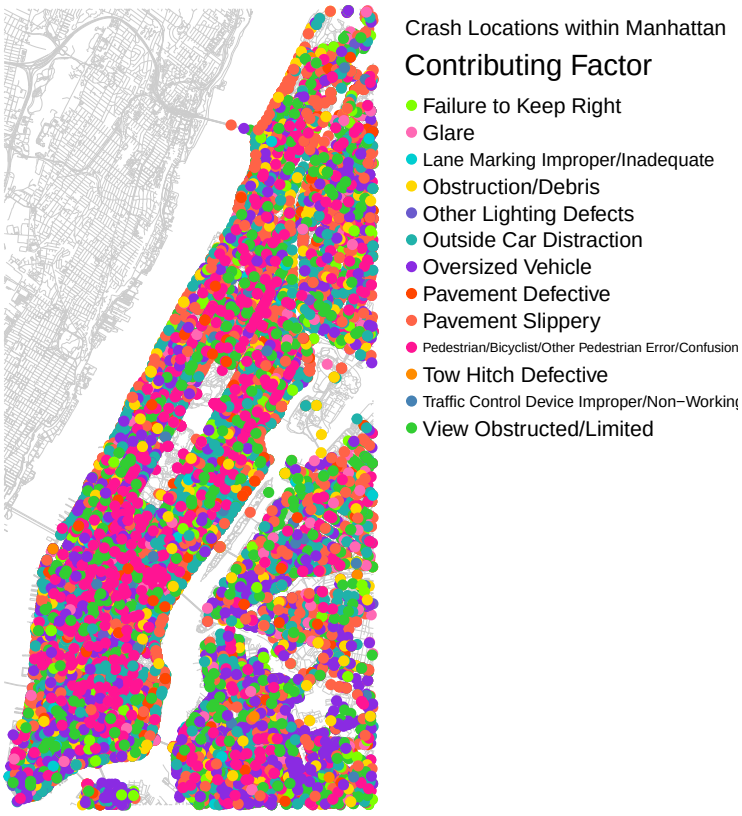

Figure 5: Analysis of Crash Locations by Contributing Factor

Analysis and Interpretation:

This plot provides a detailed visualization of bicycle crash locations across Manhattan, color-coded by contributing factors such as pavement conditions, traffic control issues, and visibility obstructions. Each colored dot represents a crash incident, with the legend indicating the specific contributing factor associated with each crash.

Key Observations:

High Density of Crashes: The plot shows a high concentration of crashes throughout Manhattan, particularly in areas like Midtown and Lower Manhattan. These regions are known for their dense traffic and complex road networks, which likely contribute to the higher incidence of crashes.

Variety of Contributing Factors: The diverse color distribution indicates that crashes are caused by a wide range of factors. For example, factors like "Pavement Slippery," "Glare," and "Lane Marking Improper/Inadequate" are frequently observed, highlighting infrastructure-related risks for cyclists.

Clusters of Specific Factors: Some areas appear to have clusters of crashes with specific contributing factors. For example, certain blocks may have multiple crashes related to "Traffic Control Device Improper/Non-Working," suggesting potential issues with traffic signal functioning or visibility in those locations.

Spatial Distribution: The crashes are widespread across the borough, indicating that while certain areas may have higher densities, the risk of crashes exists throughout Manhattan. This highlights the need for widespread safety measures rather than localized interventions.

***Interpretation*:**

The spatial distribution of crashes across Manhattan suggests that a variety of environmental and infrastructural factors contribute to cycling risks in this area. The presence of multiple contributing factors in certain areas indicates the need for targeted interventions to address specific issues like poor road conditions or inadequate traffic controls.

Moreover, the density of crashes in central areas underscores the importance of focusing on high-traffic zones for safety improvements. The data can be used by urban planners and policymakers to prioritize areas for infrastructure upgrades, such as improved pavement, better lane markings, and enhanced traffic signal visibility, to reduce the incidence of bicycle crashes.

In summary, this plot is a valuable tool for identifying risk factors and high-risk areas for bicycle crashes in Manhattan, providing a foundation for informed decision-making aimed at enhancing cyclist safety in one of the busiest boroughs of New York City.

Data Pre-Processing

In this analysis, extensive data pre-processing was performed to prepare the dataset for modeling. The steps involved include filtering bicycle-related crashes, identifying and selecting relevant factors contributing to the crashes, and spatially joining different data layers.

Given the computational intensity of these steps, the final pre-processed dataset, named 'clustering_data' , was saved for future use. This approach ensures that the modeling phase can be executed directly without the need to repeat the time-consuming pre-processing steps.

Note: To reduce computational overhead, you can skip the data pre-processing steps and load the pre-processed dataset 'clustering_data' directly into your environment for modeling.

Data Filtering and Focus on Bicycle-Related Crashes

A series of essential data pre-processing steps to prepare the dataset for subsequent analysis and modeling, particularly focusing on bicycle crashes were performed in the data pre-processing steps. Initially, keywords related to bicycles were defined and employed to filter the dataset, ensuring that only relevant records involving bicycles were retained. This filtering was done to ensure that the analysis remained centered on bicycle-related incidents. Contributing factors to crashes were identified and used to refine the dataset further, allowing a focused examination of specific road or environmental conditions that may have contributed to these crashes.

Temporal Categorization and Data Preparation for Modeling

A 'TIME.SLOT' column was added to categorize each crash by the time of day, which allowed for time-based analysis to identify peak hours for bicycle crashes. To prepare the data for modeling, categorical columns were converted into boolean values, missing data was handled by replacing it with "Not Available," and numerical data, such as speed limits, was properly formatted. The dataset was then grouped by COLLISION_ID, summarizing key features and aggregating data related to each crash. One-hot encoding was applied to categorical variables, converting them into a binary format necessary for many machine learning algorithms.

Spatial Buffer Creation and Handling Missing Values

A spatial buffer was created around each crash location to explore the immediate surroundings, including nearby amenities or road conditions that might have influenced the occurrence of crashes. Missing values were handled by removing columns with a high percentage of missing data, ensuring that the remaining data was more reliable for analysis. A spatial join was conducted to link the buffered crash locations with nearby amenities, enabling an analysis of how these amenities might have affected crash occurrences. Similarly, highways data was filtered and joined with the buffered crash locations to focus on relevant road conditions that intersected with crash sites.

Calculation of Risk Factor and Final Data Preparation

Finally, the data was prepared by combining the encoded features, removing unnecessary columns, and calculating a risk factor based on various attributes, such as the number of injuries and the time of day. This risk factor provided a quantifiable measure of crash severity, which was essential for further analysis or predictive modeling. These data pre-processing steps ensured that the dataset was clean, relevant, and optimized for the computationally intensive tasks that followed, such as clustering and modeling. Given the computational demands of these steps, the processed dataset was saved as clustering_data and was loaded directly for modeling to save time and resources.

Calculation of Risk Factor

The risk factor was calculated as a composite measure to assess the severity and potential danger associated with each bicycle crash. The calculation began by establishing a spatial buffer around each crash location, set at a 15-meter radius, to account for nearby crashes and their potential influence. The CRASH.TIME was extracted and categorized into hour segments to determine whether the crash occurred during the day or night, as time of day can significantly impact visibility and traffic conditions.

Next, the initial risk factor was set to a baseline value of 0.5, and it was further adjusted based on specific attributes of the crash. For example, the number of persons killed in the crash was multiplied by 5, reflecting the high severity associated with fatalities, while the number of persons injured was multiplied by 2.5. These weighted factors were summed to create a preliminary risk score for each crash.

The risk factor was also influenced by the time of day; crashes occurring at night were given an additional weight of 3 due to the higher risks associated with reduced visibility and potentially more dangerous driving conditions. Following this, the crash data was joined with itself to identify intersections with other nearby crashes within the buffer zone. This allowed for the inclusion of the number of nearby crashes as an additional risk component, where each intersecting crash added 2 points to the overall risk factor.

Finally, the calculated risk factors were normalized to a range between 0 and 1 to standardize the values for subsequent analysis and comparison. This normalization ensured that the risk factor could be consistently interpreted across different crashes, allowing for a clearer understanding of which incidents were associated with higher levels of danger or severity. The resulting risk factor provided a quantifiable measure that could

be used in further analysis or predictive modeling to prioritize safety interventions and understand crash dynamics better.

Overview of the Clustered Data

The clustered dataset contains 20 features and 472 fields, offering a detailed representation of various characteristics associated with bicycle crashes within a specific area. The data is organized in a simple feature collection (sf object) format, which includes geometry information, allowing for spatial analysis. Each row corresponds to a unique crash incident, identified by a COLLISION_ID, and includes attributes such as the number of persons injured or killed, the presence of street features like crossings or pedestrian signals, and whether the crash occurred on a lit or one-way street. The dataset also contains numerous categorical variables related to the types of roads involved, such as highways, residential streets, and footways, and their specific conditions like markings, surface types, and access permissions.

Spatial and Risk Analysis Features

In addition to basic crash attributes, the dataset includes a calculated risk factor for each crash, which quantifies the severity of incidents based on the number of injuries and fatalities, time of day, and proximity to other crashes. The bounding box provided indicates that the data spans a geographical area in New York City, specifically within the coordinates defined by the minimum and maximum longitude and latitude values. The inclusion of detailed spatial geometries enables the use of advanced spatial analysis techniques, allowing for a comprehensive examination of crash hotspots and the contributing factors across different types of roads and intersections. The clustering of data points based on these variables facilitates more targeted safety interventions and policy-making decisions aimed at reducing the frequency and severity of bicycle-related crashes.

Data Modeling

Explanation of the Road Risk Assessment Model

The process involves developing a Random Forest model aimed at assessing road risk based on various factors associated with bicycle crashes. The model is designed to predict a risk_factor for future crashes, which is informed by historical crash data and spatial information.

Initially, the dataset (clustering_data) is preprocessed to remove columns that are not necessary for the modeling process, such as COLLISION_ID, NUMBER.OF.PERSONS.INJURED, and NUMBER.OF.PERSONS.KILLED. The geometry data is also removed to focus on the non-spatial attributes. A seed is set for reproducibility, and the dataset is split into training (70%) and validation (30%) sets. The training and validation data are then converted back to spatial data frames (sf objects) using latitude and longitude coordinates, ensuring that spatial information is preserved for the modeling process.

Spatial Cross-Validation

The createSpatialFolds function is defined to generate spatial cross-validation folds. This function uses k-means clustering on the spatial centroids of the crash locations to divide the data into k spatial folds. This approach ensures that each fold is spatially distinct, which is important for avoiding overfitting in spatial models. The folds are then used to create indices for cross-validation within the caret package's trainControl function. By using spatial folds, the model can be evaluated in a way that more accurately reflects real-world scenarios, where crashes in different areas might have different risk factors.

Random Forest Model Training

A Random Forest model is trained using the spatial folds defined earlier. Random Forest is a robust ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification or the mean prediction for regression. In this case, the model is set up to predict the risk_factor, which represents the severity and likelihood of crashes at specific locations. The trControl parameter in the train function specifies that spatial cross-validation should be used, ensuring that the model's performance is evaluated on spatially distinct subsets of the data. The importance parameter is set to TRUE, which means that the model will calculate and return the importance of each predictor variable, providing insights into which factors contribute most to the predicted risk.

Results

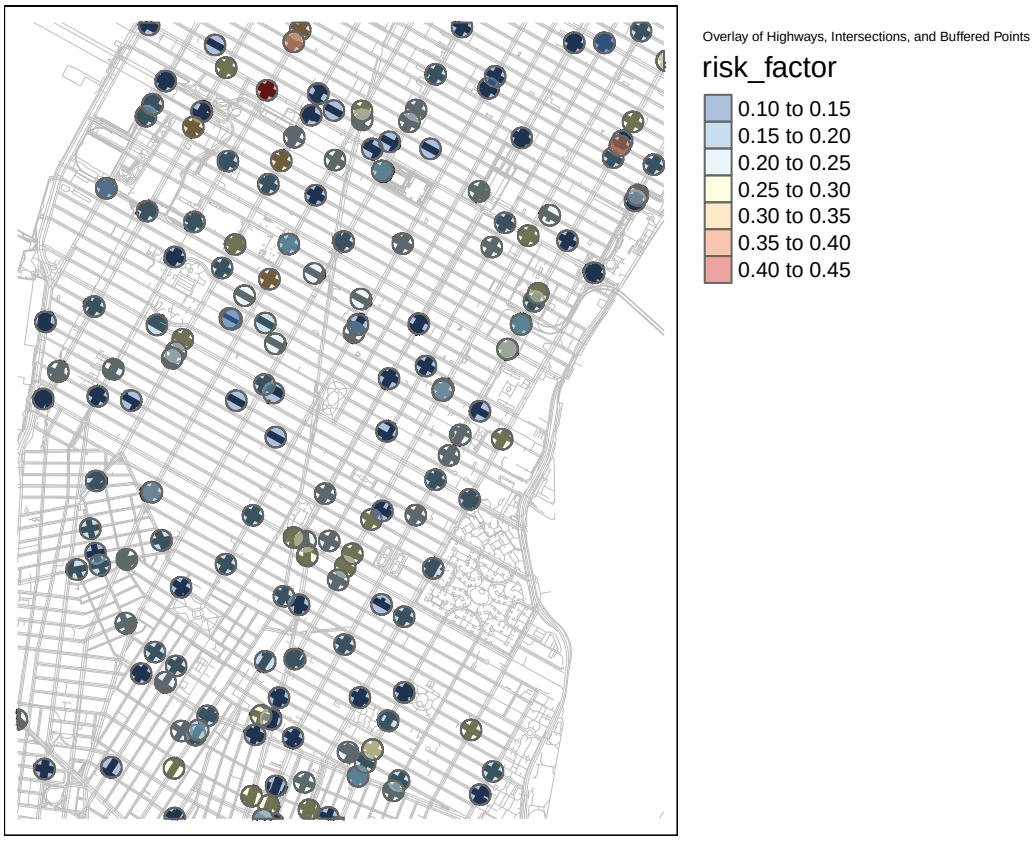

Figure 6: 1. Map Visualization of Bicycle Accident Risk Factors in Manhattan: Highlighting the intersections of highways and buffered points with varying levels of risk factors, from low (blue) to high (red)

Analysis and Interpretation:

The map presented shows an overlay of highways, intersections, and buffered points within a section of Manhattan, highlighting areas with varying levels of risk for bicyclists. The risk factor is color-coded, with darker shades indicating higher risk levels and lighter shades representing lower risk levels. This visualization is critical for understanding the spatial distribution of potential hazards for cyclists in the area.

Figure 7: 2. Map Visualization of Bicycle Accident Risk Factors in Manhattan: Highlighting the intersections of highways and buffered points with varying levels of risk factors, from low (blue) to high (red)

Risk Factor Distribution:

High-Risk Areas:

The red and orange circles indicate intersections and road segments with the highest calculated risk factors, ranging from 0.30 to 0.45. These areas are likely locations with frequent crashes, severe injuries, or fatalities. The larger size of these circles also suggests a higher concentration of accidents, which could be due to complex intersections, poor visibility, or inadequate cycling infrastructure.

Moderate to Low-Risk Areas:

The blue and yellow circles, representing risk factors between 0.05 and 0.30, show regions with moderate to lower risks. These areas might still experience accidents but with less severity or frequency compared to the high-risk zones. Factors contributing to these risk levels might include the volume of traffic, road conditions, or the presence of protective measures such as bike lanes.

Impact on Bicyclist Safety:

The plot highlights the importance of targeted interventions in specific high-risk areas to improve bicyclist safety. For instance, infrastructure enhancements like better lighting, clearer signage, and dedicated bike lanes in high-risk zones could significantly reduce the likelihood of accidents. Furthermore, this analysis can inform city planners and policymakers about the critical areas that need immediate attention to protect cyclists, ultimately contributing to safer urban mobility.

This visualization serves as a powerful tool for both understanding current risks and planning future safety measures to mitigate hazards for cyclists in Manhattan and the same can be repeated for different boroughs and parts of New York.

Results

The performance of the predictive model was evaluated based on its ability to accurately assess the risk of bicycle collisions across New York City's road network. The model's effectiveness was measured using key metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared. These metrics provide insights into how well the model's predictions align with the actual observed data, offering a comprehensive view of the model's accuracy and reliability. Model Performance Metrics

Root Mean Square Error (RMSE):

The final model achieved an RMSE of 0.05662313 for Random Forest Model and 0.0876462 for Gaussian INLA , suggesting that the model's predictions are, on average, very close to the actual risk values. This low RMSE value reflects the model's capability to consistently predict risk with minimal deviation from the observed data.

Mean Absolute Error (MAE):

The final model recorded an MAE of 0.035982 for Random Forest and 0.06052 for, which means that, on average, the model's predictions were off by a small margin. This low MAE further underscores the model's precision, indicating that the predicted risk scores are closely aligned with the actual collision risks observed in the dataset. ## R-squared:

The final model achieved an R-squared value of 0.73977 and 0.34530 for INLA, meaning that approximately 40.8% of the variability in collision risk across different locations in NYC can be explained by the features included in the model, such as road quality, lane markings, intersection density, and the presence of cycling infrastructure. While this value suggests that the model captures a significant portion of the factors influencing collision risk, it also indicates that there are additional variables or complexities that the model does not account for, which could be explored in future research.

The strong performance metrics of the model suggest that it is well-suited for identifying high-risk areas for bicycle collisions in New York City. The low RMSE and MAE values indicate that the model can accurately predict collision risks, making it a valuable tool for both cyclists and urban planners. Cyclists can use the model's predictions to choose safer routes, while city planners can identify areas that require infrastructure improvements or targeted safety interventions.

Discussion

The project provides a comprehensive analysis of road risk factors for bicyclists in New York City, with a specific focus on non-behavioral factors such as road conditions, infrastructure, and environmental variables. The analysis demonstrates that these factors significantly influence the likelihood of bicycle collisions, and by identifying high-risk areas, the study offers valuable insights for enhancing cyclist safety.

Key aspects of the project include the development of a risk factor model using Random Forest and INLA models, which were validated through spatial cross-validation techniques. These models allowed for an in-depth examination of various predictor variables, including road types, surface conditions, and time of day, and their impact on bicyclist safety. The study highlights that targeted interventions, such as improved road maintenance, better signage, and the expansion of dedicated bike lanes, could substantially reduce the risk of accidents, particularly in identified high-risk zones.

Moreover, the spatial analysis component, particularly the use of Geographic Information Systems (GIS) to visualize risk factors across Manhattan, provides a clear and actionable framework for city planners. The maps created in this study not only identify areas with the highest risk but also allow for the visualization of potential intersections where multiple risk factors converge, offering a strategic approach to implementing safety improvements.

Limitation

Despite the robustness of the analysis, there are several limitations to consider:

1. Data Quality and Completeness: The accuracy of the results is dependent on the quality and completeness of the data used. Missing or inaccurate data, particularly regarding crash reports and road conditions, could lead to biased or incomplete risk assessments.
2. Modeling Assumptions: The risk factor models developed in this study rely on several assumptions, such as the weighting of injuries and fatalities in the risk calculation. These assumptions, while based on logical reasoning, may not fully capture the complexity of real-world scenarios. Additionally, the model may not account for all potential variables that could influence crash risk, such as weather conditions or driver behavior.
3. Spatial Resolution: The spatial analysis was focused on Manhattan, which, while representative, does not encompass the entire city. Therefore, the findings may not be fully generalizable to other boroughs or regions with different road conditions and traffic patterns. The chosen buffer size and spatial resolution might also limit the granularity of the risk assessments.
4. Temporal Changes: The analysis is based on historical data and does not account for potential changes in traffic patterns, infrastructure developments, or urban policies that may have occurred since the data was collected. As a result, the risk factors identified may not reflect current or future conditions accurately.

5. Computational Intensity: The extensive data processing and model training required significant computational resources. While the study managed these challenges by focusing on a specific geographic area, scaling the analysis to cover the entire city or multiple cities would require even greater computational power, potentially limiting the feasibility of the approach for larger-scale applications.

Conclusion

The analysis presented in this project demonstrates the significant impact that non-behavioral factors, such as road conditions and infrastructure, have on the risk of bicycle collisions in New York City. By employing advanced machine learning models and spatial analysis techniques, the study offers a detailed and actionable approach to improving cyclist safety across the city.

One of the key strengths of this study lies in its ability to move beyond traditional, reactive safety measures, offering a proactive framework that anticipates risks before they manifest into accidents. By identifying high-risk areas and the underlying factors contributing to these risks, the study provides a foundation for targeted interventions. These interventions can range from infrastructure improvements, such as the installation of dedicated bike lanes, to enhanced road maintenance efforts that address issues like poor pavement conditions or inadequate signage.

Moreover, the integration of spatial analysis into the risk assessment process allows for a nuanced understanding of how different factors interact in specific geographic locations. This geographic specificity is crucial for urban planners and policymakers who need to allocate resources efficiently and prioritize interventions in areas where they will have the greatest impact on cyclist safety. The findings from this study have broader implications beyond New York City. As urban cycling continues to grow in popularity worldwide, the methodologies and insights developed here can be adapted and applied to other cities facing similar challenges. The approach provides a template for cities to assess and mitigate road risks proactively, ultimately contributing to safer and more sustainable urban environments.

However, the study also acknowledges its limitations, particularly in terms of data quality, modeling assumptions, and the focus on a specific geographic area. Addressing these limitations in future research will be essential for refining the model and ensuring that its predictions remain accurate and relevant as urban landscapes and traffic patterns evolve.

In conclusion, this project represents a significant step forward in the field of urban cycling safety. By harnessing the power of data and spatial analysis, it provides a robust tool for reducing bicycle collisions and enhancing the overall safety of urban roadways. The insights gained from this study not only have the potential to save lives in New York City but also offer valuable lessons for cities around the world as they work to promote cycling as a safe and viable mode of transportation.

References

NYC Open Data. (n.d.). Traffic Accident Reports. Retrieved from https://data.cityofnewyork.us/

Data Source Links

- 1. Motor Vehicle Collisions: Crashes - * https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data
- 2. Motor Vehicle Collisions: Vehicles - * https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Vehicles/bm4k-52h4/about_data